Arrovian Efficiency and Auditability in the Allocation of Discrete Resources*

Marek Pycia[†] M. Utku Ünver[‡]

This draft: October 2020. First draft: October 2013.

Abstract

In environments where heterogeneous indivisible resources are being allocated without monetary transfers and each agent has a unit demand, we show that an allocation mechanism is individually strategy-proof and Arrovian efficient, i.e., it always selects the best outcome with respect to some Arrovian social welfare function if, and only if, the mechanism is group strategy-proof and Pareto efficient. Re-interpreting Arrow's Independence of Irrelevant Alternatives in terms of auditability of the mechanism, we further show that these are precisely the mechanisms that are strategy-proof, Pareto efficient, and auditable.

Keywords: Individual strategy-proofness, group strategy-proofness, Pareto efficiency, Arrovian preference aggregation, matching, no-transfer allocation and exchange.

JEL classification: C78, D78

*The results in this paper were first reported in the October 2013 draft of Pycia and Ünver's "Incentive Compatible Allocation and Exchange of Discrete Resources." The present paper subsumes the analysis of Arrovian efficiency from this earlier draft. Between 2016 and 2019, this paper was posted under the title "Arrovian Efficiency in Allocation of Discrete Resources." We would like to thank Vikram Manjunath and audiences in Princeton and Montreal for their comments. Pycia gratefully acknowledges the support of UCLA, where he was a faculty member when working on this paper, as well as the financial support from the William S. Dietrich II Economic Theory Center at Princeton University.

[†]University of Zurich, Department of Economics, e-mail: `marek.pycia@econ.uzh.ch`.

[‡]Boston College, Department of Economics, e-mail: `unver@bc.edu`.

1 Introduction

Microeconomic theory has informed the design of many markets and other institutions. Many new mechanisms have been proposed to allocate resources in environments in which agents have single-unit demands and transfers are not used or are prohibited. These environments include the allocation and exchange of transplant organs, such as kidneys (Roth, Sönmez, and Ünver, 2004); the allocation of school seats in Boston, New York City, Chicago, etc. (Abdulkadiroğlu and Sönmez, 2003); and the allocation of dormitory rooms at US colleges (Abdulkadiroğlu and Sönmez, 1999). The mechanisms used elicit ordinal preferences of participants.[1]

The central concerns in the development of allocation mechanisms are incentives and efficiency.[2] The literature focused on Pareto efficiency: a matching is Pareto efficient if there exists no other matching that makes everybody weakly better off and at least one agent better off.[3] Pareto efficiency however is a weak efficiency concept; while interpersonal utility comparisons are not needed for Pareto efficiency, it only gives a lower bound for what can be achieved through desirable mechanisms. In consequence, welfare economics—starting with Bergson (1938), Samuelson (1947), and Arrow's (1963)—have long looked at stronger efficiency concepts requiring an efficient outcome to be the maximum of a social ranking of outcomes; an idea later named as resoluteness.[4] For instance, Arrow (1963), pp. 36-37, discusses the partial ordering of outcomes given by Pareto dominance, and observes:

> But though the study of maximal alternatives is possibly a useful preliminary to the analysis of particular social welfare functions, it is hard to see how any policy recommendations can be based merely on a knowledge of maximal alternatives. There is no way of deciding which maximal alternative to decide on.

Our paper is the first to carry out this program of analyzing stronger welfare criteria in

[1]In the context of deterministic mechanisms without transfers eliciting ordinal information is all we can do. In addition, eliciting ordinal preferences is considered simpler and more practical (see Bogomolnaia and Moulin, 2001).

[2]For instance, Bogolomania and Moulin (2004) discuss "a recent flurry of papers on the deterministic assignment of indivisible goods" and state that "the central question of that literature is to characterize the set of efficient and incentive compatible (strategy-proof) assignment mechanisms." The prior theoretical literature on single-unit-demand allocation without transfers has focused on characterizing mechanisms that are strategy-proof and efficient alongside other properties (see below for examples of such characterizations). In contrast, our characterization of strategy-proofness and efficiency does not rely on additional assumptions.

[3]Relatedly, constrained Pareto efficiency is also studied, e.g. stable matchings that are not Pareto dominated by other stable matchings.

[4]Resoluteness has been a standard property in social choice since its conception; in addition to Arrow's discussion see e.g. Austen-Smith and Banks (1999) for the role of resoluteness in political science, and Zwicker (2016) for a recent survey of canonical social choice results such as Gibbard (1973)-Satterthwaite (1975) Theorem that implicitly or explicitly involve resoluteness. The failure of resoluteness is also at the core of the Condorcet paradox, see e.g. Black (1948) and Campbell and Kelly (2003).

the context of allocation of discrete resources without transfers.[5] We analyze both social choice functions as well as social welfare functions satisfying Arrow's postulates. A social choice function (SCF) determines what unique alternative should be implemented for each preference profile, while social welfare function (SWF) determines a societal ranking of outcomes for each profile of individual preference rankings. Allowing for partial societal rankings, we can treat an SCF as an SWF in which the outcome of SCF is ranked above all other outcomes.[6] Following Arrow (1963), we say that an SWF is Arrovian if, and only if, it satisfies the standard resoluteness, Pareto, and independence-of-irrelevant-alternatives postulates. An SWF is resolute if it has a unique social maximum for each profile of preferences; in particular, every SCF is resolute. An SWF satisfies the Pareto postulate if two socially and Pareto-comparable matchings are ranked so that the Pareto-dominant matching is ranked above the Pareto-dominated one. An SWF satisfies the independence of irrelevant alternatives if, given any two profiles of preferences and any two matchings that are socially comparable under both profiles, if all agents rank the two matchings in the same way under both profiles, then the social ranking of the two matchings is the same under both profiles. When we want to highlight the positive rather than normative aspects of an SCF we refer to it as a mechanism; we allow here both Arrovian and not Arrovian SCFs. We call a mechanism efficient with respect to an SWF if, for every preference profiles, the resulting matching is a maximum of the SWF.[7] We say that a mechanism is Arrovian efficient if it is efficient with respect to some Arrovian SWF. Finally, we say that a mechanism is strategy-proof if, for any reports by other agents, reporting the true ranking leads to the mechanism outcome being weakly better for an agent than any other report.

We provide a new motivation for Arrovian efficiency by showing that it is equivalent to a mild auditability requirement that in order to falsify a proposed mechanism outcome, it is sufficient to verify pairwise comparison of agents' preferences of the outcome with only one alternative (challenger) outcome. This auditability property is attractive as it allows to falsify the mechanism outcome with a limited amount of information and thus largely preserves the privacy of participants' information. Restricting attention to strategy-proof mechanisms, we also show that Arrovian independence of irrelevant alternatives is equivalent

[5]We study the canonical environment with finite numbers of agents and indivisible objects, dubbed as "houses" (Shapley and Scarf, 1974). Each agent has a strict preference relation over objects. An outcome—or a matching—specifies for each agent an object the agent is matched with in such a way that no object is matched with two different agents.

[6]For analysis of welfare with partial orderings, see e.g. see Sen (1970,1999) and Weymark (1984).

[7]There is a rich social choice literature on the correspondence between choice and the maximum of the SWF ranking in the context of social choice (see below). This literature is interested in rationalizing social choice rather than the efficiency of allocation mechanisms, and hence it talks about mechanisms "rationalized by an SWF" rather than "efficient with respect to an SWF."

to non-bossiness of Satterthwaite and Sonnenschein (1981), which allows us to leverage the results of Pycia and Ünver (2017) to fully characterize the class of auditable and efficient mechanisms as the class of Trading Cycles mechanisms. This characterization provides a no-transfer counterpart of Akbarpour and Li (2020) insight that classical auctions are the "credible" mechanisms in their sense.[8]

We use this characterization to show that almost sequential dictatorships are the only mechanisms that are individually strategy-proof and Arrovian efficient with respect to an SWF that ranks all matchings. An almost sequential dictatorship combines the ideas of sequential dictatorship and majority voting between only two possible outcomes. Dictatorships are the benchmark strategy-proof and efficient mechanisms in many areas of economics. For instance, Gibbard (1973) and Satterthwaite (1975) have shown that all strategy-proof and unanimous voting rules are dictatorial. Moreover, for this result to hold we need more than two alternatives. With two alternatives there are other mechanisms that are strategy-proof and unanimous (majority voting being the primary example), very much like our class of almost sequential dictatorships.[9] Still, we find it surprising that this theorem is true in our environment because — in contrast to the environments where this question was previously studied — ours allows many individually strategy-proof (and even group strategy-proof) and Pareto-efficient mechanisms that are not dictatorial.

The present paper is the first to connect the literature on allocation and exchange of discrete resources and the literature on Arrovian preference aggregation. In particular, we seem to be the first to recognize the equivalence of Theorem 1. However, stronger equivalence results — which do not hold true in our setting — are familiar from studies of voting. In voting — unlike in our problem — all agents have strict preferences among all outcomes. In the class of Pareto-efficient mechanisms, individual strategy-proofness is then equivalent to group strategy-proofness (Gibbard, 1973, and Satterthwaite, 1975).[10] This stronger equivalence fails in our setting, as it admits individually strategy-proof and Pareto-efficient

[8]See also Woodward (2020) for an analysis of a more general concept of auditability in multi-unit auctions. For the literature on privacy in mechanism design see the recent survey Pai and Roth (2018).

[9]Dasgupta, Hammond, and Maskin (1979) extended this result to more general social choice models, Satterthwaite and Sonnenschein (1981) extended it to public goods economies with production, Zhou (1991) extended it to pure public goods economies. In exchange economies, Barberà and Jackson (1995) showed that strategy-proof mechanisms are Pareto inefficient.

[10]The equivalence of Theorem 1 has counterparts in the social choice literature on restricted preference domains—such as single-peaked preferences—in which there are non-dictatorial strategy-proof and Arrow efficient rules. For instance, Moulin (1988) extends a result by Blair and Muller (1983) and shows that in environments such as single-peaked voting, if an Arrovian SWF is monotonic, then the mechanism picking its unique maximal element is group strategy-proof. In particular, this implies that in single-peaked voting individual strategy-proofness and group strategy-proofness are equivalent, with no need to restrict attention to efficient mechanisms. In contrast, in allocation environments the equivalence results from the conjunction of incentive and efficiency assumptions, and the equivalence of incentive assumptions alone is not true.

mechanisms that fail group strategy-proofness. As far as we know, this is also the first paper that provides a joint approach to mechanism design and Arrovian preference aggregation.

Our paper also contributes to the literature on characterizations of dominant strategy mechanisms for house allocation. Ehlers (2002) characterizes group-strategy-proof and Pareto-efficient mechanisms in a maximal domain of weak preferences for which such mechanisms exist and proves a general impossibility result for the domain of all weak preferences.[11] Note that our concept of partial social ranking is different from Ehlers' allowing only certain weak preferences over assigned houses; Ehlers' work is not concerned with social rankings of outcomes. Pycia and Ünver (2017) characterizes group-strategy-proof and Pareto-efficient mechanisms in the standard domain of strict preferences and Root and Ahn (2020) characterize properties of these mechanisms allowing for constraints and providing a synthetic treatment of many social choice domains; see also Barberà (1983) and Pápai (2000) who laid the foundations for this line of research. Ma (1994) characterized the class of strategy-proof, individually rational, and Pareto-efficient mechanisms, and his characterization has been extended by Pycia and Ünver (2017) and Tang and Zhang (2015) to richer single-unit demand, by Pápai (2007) to multi-unit demand models, and by Pycia (2016) to settings with network constraints.[12]

At the heart of the above characterizations is David Gale's idea of Top Trading Cycle (TTC) mechanism, first reported in Shapley and Scarf (1974). This mechanism was extended beyond the house exchange domain of Shapley and Scarf (1974) by Abdulkadiroğlu and Sönmez (1999) (allowing for a mixture of allocation and exchange, see also Sönmez and Ünver, 2010), Abdulkadiroğlu and Sönmez (2003) (allowing for copies), Pápai (2000), Pycia and Ünver (2017), Pycia and Ünver (2011) (making the mechanism more flexible by allowing for richer classes of property rights), Jaramillo and Manjunath (2012) (allowing for weak preferences), Hakimov and Kesten (2014) and Morrill (2015) (making the mechanism more equitable), and Pycia (2016) (allowing for constraints).

Sequential dictatorships have not been studied extensively with unit demand for goods, although their special cases have been. In a *serial dictatorship* (also known as a *priority*

[11]Most of the literature on house allocation—including our paper—is not affected by Ehlers' impossibility result because it analyzes environments in which agents' preferences are strict.

[12]Other contributions characterize dominant strategy and efficient mechanisms that satisfy selected additional assumptions include Ehlers, Klaus, and Pápai (2002) and Ehlers and Klaus (2003) who characterized strategy-proof mechanisms with population and resource monotonicity properties, respectively. Karakaya, Klaus, and Schlegel (2019) replaced group strategy-proofness by individual strategy-proofness together with other properties such as consistency. Abdulkadiroğlu, Che, Pathak, Roth, and Tercieux (2020) characterized strategy-proofness and a minimum envy generalization of individual rationality. Characterization results with multiple copies of objects were obtained by Liu and Pycia (2011), Morrill (2013), Pycia (2019), and Pycia and Troyan (2019). The core mechanism in more complex exchange markets was characterized by Pápai (2007).

mechanism), the same agent chooses next regardless of which house the current agent picks. Svensson (1994) formally introduced and studied serial dictatorships first; Abdulkadiroğlu and Sönmez (1998) studied a probabilistic version of them where the order of agents is determined uniformly randomly; Svensson (1999) and Ergin (2000) characterized them using plausible axioms. Allowing for outside options, Pycia and Ünver (2007) characterized a subclass of sequential dictatorships different from serial dictatorships. With multiple-house demand under responsive preferences, Hatfield (2009) showed that sequential dictatorships are the only strategy-proof, non-bossy, and Pareto-efficient mechanisms, and Pápai (2001) characterized the sequential dictatorships through the properties of strategy-proofness, non-bossiness, and citizen sovereignty (see also Klaus and Miyagawa, 2002). In a general model allowing both the cases with and without transfers, Pycia and Troyan (2019) showed that a broad class closely resembling sequential dictatorships are precisely the mechanisms that are strongly obviously strategy-proof in their sense; see also Li (2015) and Pycia (2019). For characterizations of random serial dictatorships in terms of incentives, efficiency, and fairness see Liu and Pycia (2011) and Pycia and Troyan (2019). Root and Ahn (2020) characterize the constrained social choice domains in which generalized sequential dictatorships are the only group strategy-proof and Pareto-efficient mechanisms. As an application of their general theorem, they characterize sequential dictatorships as the only mechanisms which are group strategy-proof and Pareto efficient in the roommates problem.

2 Model

2.1 House Allocation Problems

Let I be a set of **agents** and H be a set of **objects** that we often refer to as **houses**, following the standard terminology of the literature. We use letters i, j, k to refer to agents and h, g, e to refer to houses. Each agent i has a **strict preference relation** over H, i.e., a complete, anti-symmetric, and transitive binary relation, denoted by $\succ_i$.[13] Let $\mathbf{P}_i$ be the set of strict preference relations for agent i, and let $\mathbf{P}_J$ denote the Cartesian product $\times_{i\in J}\mathbf{P}_i$ for any $J \subseteq I$. Any profile $\succ = (\succ_i)_{i\in I}$ from $\mathbf{P} \equiv \mathbf{P}_I$ is called a **preference profile**. For every $\succ \in \mathbf{P}$ and $J \subseteq I$, let $\succ_J = (\succ_i)_{i\in J} \in P_J$ be the restriction of $\succ$ to J. A **house allocation problem** is the triple $\langle I, H, \succ\rangle$ (see Hylland and Zeckhauser, 1979).

Throughout the paper, we fix I and H, and thus a problem is identified with its preference profile. We follow the tradition adopted by many papers in the literature (e.g., Svensson,

[13]We denote its induced weak-preference relation by $\succeq_i$, that is, for any $g, h \in H$, $g \succeq_i h \iff g = h$ or $g \succ_i h$.

1999) and assume that $|H| \geq |I|$ so that each agent is allocated a house.

An outcome of a house allocation problem is a matching. To define a matching, let us start with a more general concept that we will use frequently. A **submatching** is an allocation of a subset of houses to a subset of agents, such that no two different agents get the same house. Formally, a submatching is a one-to-one function $\sigma : J \to H$; where for $J \subseteq I$, using the standard function notation, we denote by $\sigma(i)$ the assignment of agent $i \in J$ under σ, and by $\sigma^{-1}(h)$ the agent that got house $h \in \sigma(J)$ under σ. Let $\mathcal{S}$ be the set of submatchings. For each $\sigma \in \mathcal{S}$, let I_σ denote the set of agents matched by σ and $H_\sigma \subseteq H$ denote the set of houses matched by σ. For every $h \in H$, let $\mathcal{S}_{-h} \subset \mathcal{S}$ be the set of submatchings $\sigma \in \mathcal{S}$ such that $h \in H - H_\sigma$, i.e., the set of submatchings at which house h is unmatched. By virtue of the set-theoretic interpretation of functions, submatchings are sets of agent-house pairs and are ordered by inclusion. A **matching** is a maximal submatching; that is, $\mu \in \mathcal{S}$ is a matching if $I_\mu = I$. Let $\mathcal{M} \subset \mathcal{S}$ be the set of matchings. We will write $\overline{I_\sigma}$ for $I - I_\sigma$ and $\overline{H_\sigma}$ for $H - H_\sigma$ for short. We will also write $\overline{\mathcal{M}}$ for $\mathcal{S} - \mathcal{M}$.

A **mechanism** or a **social choice function** is a mapping $\varphi : \mathbf{P} \to \mathcal{M}$ that assigns a matching for each preference profile (or, equivalently, for each allocation problem).[14] Denoting by $P^{\mathcal{M}}$ the set of strict partial orderings over matchings, where a strict partial ordering is a binary relation that is anti-symmetric and transitive. We refer to elements of $P^{\mathcal{M}}$ as **social rankings**. A **social welfare function** (**SWF**) $\Phi : \mathbf{P} \to P^{\mathcal{M}}$ maps agents' preference profiles to strict social rankings. If a matching μ is ranked higher than some other matching ν under $\Phi(\succ)$, we denote this as $\mu\ \Phi(\succ)\ \nu$. An SWF Φ is **resolute** if: for every preference profile $\succ$ there exists a matching $\mu \in \mathcal{M}$ such that $\mu\ \Phi(\succ)\ \nu$ for every $\nu \in \mathcal{M} - \{\mu\}$. We assume SWFs we consider are resolute. A mechanism can be identified with a special instance of a resolute SWF in which the mechanism outcome is the unique maximal outcome of the SWF and no comparisons between non-maximal outcomes are made.

2.2 Efficiency, Auditability, and Strategy-Proofness

A matching is Pareto efficient if no other matching would make everybody weakly better off and at least one agent better off. That is, a matching $\mu \in \mathcal{M}$ is **Pareto efficient** if there exists no matching $\nu \in \mathcal{M}$ such that for every $i \in I$, $\nu(i) \succeq_i \mu(i)$, and for some $i \in I$, $\nu(i) \succ_i \mu(i)$. An SWF Φ satisfies the **Pareto** postulate (or is **unanimous**) if: for every preference profile $\succ$ and any two matchings $\mu, \nu \in \mathcal{M}$ that are comparable by $\Phi(\succ)$, if $\mu(i) \succeq_i \nu(i)$ for every $i \in I$, with at least one strict preference, then $\mu\ \Phi(\succ)\ \nu$. In particular, a mechanism is **Pareto efficient** if it finds a Pareto-efficient matching for every problem.

[14] We study direct mechanisms.

Pareto efficiency is a weak efficiency requirement and, as discussed in the introduction, Arrow criticized it for its failure to uniquely determine the best outcome; that is for not being resolute.[15]

An SWF Φ satisfies the **independence of irrelevant alternatives (IIA)** if: for every $\succ, \succ' \in \mathbf{P}$ and $\mu, \nu \in \mathcal{M}$, if all agents rank μ and ν in the same way and both $\Phi(\succ)$ and $\Phi(\succ')$ rank μ and ν, then $\mu \; \Phi(\succ') \; \nu \iff \mu \; \Phi(\succ) \; \nu$.

We say that a matching μ is **efficient with respect to a social ranking** $\succ^{\mathcal{M}} \in P^{\mathcal{M}}$ if it maximizes the social welfare, that is $\mu \succ^{\mathcal{M}} \nu$ for every $\nu \in \mathcal{M} - \{\mu\}$. A mechanism φ is **efficient with respect to an SWF** Φ if for any profile of agents' preferences $\succ$, the matching $\varphi[\succ]$ is efficient with respect to $\Phi(\succ)$. If φ is efficient with respect to some SWF that satisfies the Arrovian postulates of resoluteness, Pareto, and IIA, then we say that φ is **Arrovian efficient**. The next section offers two examples illustrating the concept of Arrovian efficiency.

A mechanism φ is **auditable** (or **one-comparison auditable**) if for any preference profile $\succ$ and any matching $\nu \neq \varphi[\succ]$ and any other preference profile $\succ'$ such that the comparisons of ν and $\varphi[\succ]$ are the same under $\succ$ and $\succ'$, we have $\varphi[\succ'] \neq \nu$. This concept captures the idea that, in order to falsify a proposed matching as being the outcome of the mechanism, it is sufficient to find one challenger outcome and to verify the pairwise comparisons of the proposed outcome with the challenger. We can thus falsify an outcome with a limited amount of information; one of the reasons this is an attractive feature of a mechanism is that it allows challenges that rely on relatively little information and largely preserve agents' privacy.

A mechanism is individually strategy-proof if truthful revelation of preferences is a weakly dominant strategy for any agent: a mechanism φ is **individually strategy-proof** if for every $\succ \in \mathbf{P}$, there is no $i \in I$ and $\succ'_i \in \mathbf{P}_i$ such that

$$\varphi[\succ'_i, \succ_{-i}](i) \;\succ_i\; \varphi[\succ](i).$$

A mechanism is group strategy-proof if there is no group of agents that can misstate their preferences in a way such that each one in the group gets a weakly better house and at least one agent in the group gets a better house, irrespective of the preference ranking of the agents not in the group. Formally, a mechanism φ is **group strategy-proof** if for every

[15]In particular, when imposed on group strategy-proof mechanisms (defined below), Pareto efficiency is equivalent to assuming that the mechanism maps $\mathbf{P}$ onto the entire set of matchings $\mathcal{M}$. This surjectivity property is known as **citizen sovereignty**, or full range. Notice also that Pareto dominance is a non-resolute SWF.

$\succ \in \mathbf{P}$, there exists no $J \subseteq I$ and $\succ'_J \in \mathbf{P}_J$ such that

$$\varphi[\succ'_J, \succ_{-J}](i) \ \succeq_i \ \varphi[\succ](i) \text{ for every } i \in J,$$

and

$$\varphi[\succ'_J, \succ_{-J}](j) \succ_j \varphi[\succ](j) \text{ for at least one } j \in J.$$

3 Equivalence

In Theorem 1, we study individually strategy-proof and Pareto-efficient mechanisms and establish for them equivalence between Arrovian efficiency, auditability, and group strategy-proofness.[16] Thus, the three resulting classes of mechanisms coincide, and in particular, each class consists of Trading Cycle mechanisms of Pycia and Ünver (2017), who constructed the class of group strategy-proof and Pareto-efficient mechanisms. In addition, Example 1 below demonstrates that the class of individually strategy-proof and Pareto-efficient mechanisms is a strict superset of the mechanisms satisfying any of the equivalent conditions of the theorem.

Theorem 1. *Suppose that a mechanism is individually strategy-proof and Pareto efficient. Then the following three conditions are equivalent for this mechanism: Arrovian efficiency, auditability, and group strategy-proofness.*

Our proof establishes that some of the implications in this theorem are satisfied without restricting attention to strategy-proof and Pareto-efficient mechanisms:

Proposition 1. *If a mechanism is Arrovian efficient, then it is Pareto efficient, non-bossy, and auditable.*

To illustrate the equivalence of the theorem and our concepts, let us look at the setting with three agents 1, 2, and 3, three houses A, B, and C, and no outside options. In the Appendix, we give an example of a more elaborate incomplete Arrovian SWF, here let us consider two examples of mechanisms illustrating the conditions we study.

Example 1. The serial dictatorship φ in which 1 chooses first and 2 chooses second is well-known to be group strategy-proof and Pareto efficient. It is straightforward to see that this serial dictatorship is Arrovian efficient with respect to the following SWF: μ is ranked above ν if and only if (a) 1 prefers μ to ν, or (b) 1 is indifferent and 2 prefers μ to ν.

[16] In fact, our proof shows something more: for the mechanisms we study, auditability (or group strategy-proofness) is also equivalent to Arrovian efficiency with respect to an SWF in which if matching μ Pareto dominates matching μ' then these two matchings are comparable.

As φ treats all objects in a symmetric (neutral) way, to establish the serial dictatorship's auditability, it is sufficient to look at a preference profile $\succ$ such that $\varphi[\succ] = \{(1, A), (2, B), (3, C)\}$, a different matching ν and any preference profile $\succ'$ such that $\succ'_i$ keeps the same ranking as $\succ$ between $\varphi[\succ]$ and ν for each agent i and to show that $\varphi[\succ'] \neq \nu$. To verify this inequality consider two cases:

- $A \neq \nu(1)$. Then $A \succ_1 \nu(1)$ because 1 being the first dictator chose her top choice under $\succ_i$. Hence, $A \succ'_1 \nu(1)$. 1 is not choosing $\nu(1)$ when having preference ranking $\succ'_1$ and thus $\varphi[\succ'] \neq \nu$.
- $A = \nu(1)$. Then either :
  - $\star$ $B \neq \nu(2)$. Then, $B \succ_2 \nu(2)$ by an argument similar to the previous case. If $\varphi[\succ'](1) = B$ then $\varphi[\succ'] \neq \nu$, and the auditability inequality obtains. If $\varphi[\succ'](1) \neq B$ then either $\varphi[\succ'](1) \neq A = \nu(1)$ and the auditability inequality obtains, or $\varphi[\succ'](1) = A = \nu(1)$ and hence B is available when 2's assignment is determined, and thus, $\varphi[\succ'] \succsim'_2 B \succ'_2 \nu(2)$, and hence, $\varphi[\succ'] \neq \nu$ and the auditability inequality obtains.
  - $\star$ $B = \nu(2)$. Then $C = \nu(3)$ contrary to $\nu \neq \varphi[\succ]$.

Example 2. We now modify the serial dictatorship of the previous example and consider mechanisms ψ in which 1 chooses first; then 2 chooses second if 1 prefers B over C, else 3 chooses second. This mechanism is an example of a ranking-dependent sequential dictatorship, and is also individually strategy-proof and Pareto efficient. However, mechanism ψ is neither Arrovian efficient nor group strategy-proof nor auditable. To see the latter three points, let us look at the following two preference profiles, which differ only in how agent 1 ranks objects:

$$\succ = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & A & A \\ B & B & B \\ C & C & C \end{array} \qquad \succ' = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & A & A \\ C & B & B \\ B & C & C \end{array},$$

and notice that

$$\begin{aligned} \psi[\succ] &= \{(1, A), (2, B), (3, C)\}, \\ \psi[\succ'] &= \{(1, A), (2, C), (3, B)\}. \end{aligned}$$

Mechanism ψ fails group strategy-proofness because the coalition $\{1, 3\}$ can improve by reporting $\succ'_{\{1,3\}}$ instead of $\succ_{\{1,3\}}$. Mechanism ψ also fails Arrovian efficiency. Indeed, by way of contradiction assume that ψ is Arrovian efficient with respect to some Arrovian SWF Ψ. Then $\Psi(\succ)$ ranks allocation $\psi[\succ]$ above $\psi[\succ']$, and $\Psi(\succ')$ ranks $\psi[\succ']$ above $\psi[\succ]$. But, this violates IIA, a contradiction that shows that ψ is not Arrovian efficient. Mechanism ψ also fails auditability as we can contest the matching $\psi[\succ]$ with matching $\nu = \psi[\succ']$.

The proof of Theorem 1 builds on Example 1. As preparation for the proof, let us notice three properties of group strategy-proofness. First, in the environment we study, group strategy-proofness is equivalent to the conjunction of two non-cooperative properties: individual strategy-proofness and non-bossiness.[17] Non-bossiness (Satterthwaite and Sonnenschein, 1981) means that no agent can misreport her preferences in such a way that her allocation is not changed but the allocation of some other agent is changed: a mechanism φ is **non-bossy** if for every $\succ\in \mathbf{P}$, there is no $i \in I$ and $\succ_i'\in \mathbf{P}_i$ such that

$$\varphi[\succ_i', \succ_{-i}](i) = \varphi[\succ](i) \quad \text{and} \quad \varphi[\succ_i', \succ_{-i}] \neq \varphi[\succ].$$

The following lemma is due to Pápai (2000):

Lemma 1. *Pápai (2000) A mechanism is group strategy-proof if and only if it is individually strategy-proof and non-bossy.*

Second, in the environment we study group strategy-proofness is equivalent to Maskin monotonicity (Maskin, 1999). A mechanism φ is **Maskin monotonic** if $\varphi[\succ'] = \varphi[\succ]$ whenever $\succ'\in \mathbf{P}$ is a φ-monotonic transformation of $\succ\in \mathbf{P}$. A preference profile $\succ'\in \mathbf{P}$ is a **φ-monotonic transformation** of $\succ\in \mathbf{P}$ if

$$\{h \in H : h \succeq_i \varphi[\succ](i)\} \supseteq \{h \in H : h \succeq_i' \varphi[\succ](i)\} \text{ for every } i \in I.$$

Thus, for each agent, the set of houses better than the base-profile allocation weakly shrinks when we go from the base profile to its monotonic transformation. The following lemma was proven by Takamiya (2001) for a subset of the problems we study; his proof can be extended to our more general setting.

Lemma 2. *A mechanism is group strategy-proof if and only if it is Maskin monotonic.*

Finally, let us notice the following:

Lemma 3. *If a mechanism φ is group strategy-proof, then no agent can change the outcome of φ by changing the ranking of houses worse than the house she obtains; that is, if $\succ'$ differs from $\succ$ only in how some agent i ranks houses below $\varphi[\succ](i)$, then $\varphi[\succ'] = \varphi[\succ]$.*

[17]Both of these properties are non-cooperative in the sense that they relate a mechanism's outcomes under two scenarios when a single agent makes unilateral preference-revelation deviations.

The proof is straightforward: by individual strategy-proofness $\varphi[\succ](i) = \varphi[\succ'](i)$ and hence by group strategy-proofness (or by non-bossiness) $\varphi[\succ'] = \varphi[\succ]$.

Proof of Theorem 1 and Proposition 1. We start by establishing the first equivalence of the theorem, between Arrovian efficiency and group strategy-proofness. In proving this result, we also establish the proof of the proposition. First, consider an Arrovian efficient mechanism φ with respect to some SWF Φ. In light of Lemma 1, to establish the first implication of the equivalence as well as the comment following the theorem, it is enough to show that φ is Pareto efficient and non-bossy.

To show that φ is Pareto efficient, suppose that for some $\succ \in \mathbf{P}$, $\varphi[\succ]$ is not Pareto efficient. Then there exists some $\mu \in \mathcal{M} - \{\varphi[\succ]\}$ such that $\mu(i) \succeq_i \varphi[\succ](i)$ for every i, with a strict preference for at least one agent. Because Φ satisfies the Pareto postulate, we have $\mu\, \Phi(\succ)\, \varphi[\succ]$, which contradicts the assumption that φ is Arrovian efficient with respect to Φ.

To show that φ is non-bossy, let $\succ \in \mathbf{P}$ and $\succ_i' \in \mathbf{P}_i$ be such that

$$\varphi[\succ](i) = \varphi[\succ_i', \succ_{-i}](i).$$

Denote $\succ' = (\succ_i', \succ_{-i})$. Because φ is Arrovian efficient with respect to Φ, the matching $\varphi[\succ]$ is ranked as the unique first by $\Phi(\succ)$ and the matching $\varphi[\succ']$ is ranked as the unique first by $\Phi(\succ')$. Thus, $\varphi[\succ]$ and $\varphi[\succ']$ are comparable under both $\Phi(\succ)$ and $\Phi(\succ')$, and IIA implies that $\varphi[\succ]$ and $\varphi[\succ']$ are ranked in the same way by $\Phi(\succ)$ and $\Phi(\succ')$. We, thus, conclude that $\varphi[\succ] = \varphi[\succ']$. This establishes that φ is non-bossy.

Second, consider a group strategy-proof and Pareto-efficient mechanism φ. We define the SWF Φ as follows: for any profile of preferences $\succ$ and any matchings μ and $\mu' \neq \mu$, matching μ is ranked by $\Phi(\succ)$ above μ' iff either (i) we have $\mu = \varphi[\succ]$ or (ii) for every agents i, we have $\mu(i) \succeq_i \mu'(i)$. Note that Pareto efficiency of φ implies that conditions (i) and (ii) are consistent with each other, and hence, that the SWF Φ is well-defined.

By definition, Φ satisfies the Pareto postulate. Furthermore, Φ is transitive: if $\Phi(\succ)$ ranks μ^1 above μ^2 and it ranks μ^2 above μ^3 , then it ranks μ^1 above μ^3. Indeed, if one of the these (for $\ell = 1, 2, 3$) equals $\varphi[\succ]$, then it must be that $\mu^1 = \varphi[\succ]$, and the claim is proven. If none of the μ^i equals $\varphi[\succ]$, then agents unanimously rank μ^1 above μ^2 and unanimously rank μ^2 above μ^3; we can conclude that the agents unanimously rank μ^1 above μ^3, and thus, $\Phi(\succ)$ ranks μ^1 above μ^3.

It remains to check that Φ satisfies IIA. Take two preference profiles $\succ^1$ and $\succ^2$ such that each agent ranks two matchings, say μ and μ', in the same way under the two preference profiles. If the two matchings are comparable under both $\Phi(\succ^1)$ and $\Phi(\succ^2)$, then one of the following cases obtains:

Case 1: One of the matchings is unanimously preferred to the other under $\succ^1$; then the same unanimous preference obtains under $\succ^2$ and the claim is true.

Case 2: There is no unanimous ranking of the two matchings under $\succ^1$; then unanimity cannot obtain under $\succ^2$ either. As the matchings are ranked, it must be that $\varphi[\succ^1]$ and $\varphi[\succ^2]$ take value in $\{\mu, \mu'\}$. Say, $\varphi[\succ^1] = \mu$; then we need to check that $\varphi[\succ^2] = \mu$ as well. By Lemma 2, we can assume that each agent i ranks $\mu(i)$ and $\mu'(i)$ at the top of her ranking under both $\succ^1$ and $\succ^2$. Furthermore, by Lemma 3, only rankings of houses above agents' allocations (and including their allocations) affect the outcome of a group strategy-proof mechanism; we can thus conclude that $\varphi[\succ^1] = \varphi[\succ^2]$.

To complete the proof, we need to also establish the equivalence between auditability and the other two concepts. An inspection of the definitions shows that Arrovian efficiency directly implies auditability; indeed, auditability is effectively IIA restricted to comparisons involving the top alternative. Second, notice that in the proof of non-bossiness we only relied on such restricted IIA. Hence, auditability implies non-bossiness and Lemma 1 concludes the proof. QED

For our next section when we consider complete SWFs, we need to introduce the full class of group-strategy-proof and Pareto-efficient mechanisms, as characterized by Pycia and Ünver (2017). This is the class of trading-cycles mechanisms. This mechanism class is defined through an iterative algorithm, which matches some agents in every round. Depending on who is matched with which house in preceding rounds, the remaining houses are controlled by the remaining agents in a round of the algorithm. We define a control-rights structure as a function of the submatching that is fixed:

Definition 1. A **structure of control rights** is a collection of mappings

$$(c, b) = \left\{(c_\sigma, b_\sigma) : \overline{H_\sigma} \to \overline{I_\sigma} \times \{\text{ownership,brokerage}\}\right\}_{\sigma\in\overline{\mathcal{M}}}.$$

The functions c_σ of the control-rights structure tell us which unmatched agent controls any particular unmatched house at a submatching σ, where **at** σ is the terminology we use when some agents and houses are already matched with respect to σ. Agent i **controls** house $h \in \overline{H_\sigma}$ at submatching σ when $c_\sigma(h) = i$. The type of control is determined by functions b_σ. We say that the agent $c_\sigma(h)$ **owns** h at σ if $b_\sigma(h)$ =ownership, and that the agent $c_\sigma(h)$ **brokers** h at σ if $b_\sigma(h)$ =brokerage. In the former case, we call the agent an **owner** and the controlled house an **owned house**. In the latter case, we use the terms **broker** and **brokered house**. Notice that each controlled (owned or brokered) house is unmatched at σ, and any unmatched house is controlled by some uniquely determined unmatched agent.

We need to impose certain conditions on the control-rights structures to guarantee that the induced mechanisms are group strategy-proof and Pareto efficient.

Definition 2. A structure of control rights (c, b) is **consistent** if the following within-round and across-round requirements are satisfied *for every* $\sigma \in \overline{\mathcal{M}}$:

Within-Round Requirements:

> *(**R1**) There is at most one brokered house at σ, or $|\overline{H_\sigma}| = 3$ and all remaining houses are brokered.*
>
> *(**R2**) If i is the only unmatched agent at σ, then i owns all unmatched houses at σ.*
>
> *(**R3**) If agent i brokers a house at σ, then i does not control any other houses at σ.*

Across-Round Requirements: *Consider submatching σ' such that $\sigma \subset \sigma' \in \overline{\mathcal{M}}$, and an agent $i \in \overline{I_{\sigma'}}$ that owns a house $h \in \overline{H_{\sigma'}}$ at σ. Then:*

> *(**R4**) Agent i owns h at σ'.*
>
> *(**R5**) If i' brokers house h' at σ, and $i' \in \overline{I_{\sigma'}}$, $h' \in \overline{H_{\sigma'}}$, then either i' brokers h' at σ', or i owns h' at σ'. (Notice that the latter case can only happen if i is the only agent in $\overline{I_{\sigma'}}$ who owns a house at σ.)*
>
> *(**R6**) If agent $i' \in \overline{I_{\sigma'}}$ controls $h' \in \overline{H_{\sigma'}}$ at σ, then i' owns h at $\sigma \cup \{(i, h')\}$.*

Each consistent control-rights structure (c, b) induces a **trading-cycles (TC)** mechanism $\psi^{c,b}$, and given a problem $\succ \in \mathbf{P}$, the outcome matching $\psi^{c,b}[\succ]$ is found as follows:

> **The TC algorithm.** The algorithm starts with empty submatching $\sigma^0 = \varnothing$ and in each round $r = 1, 2, ...$ it matches some agents with houses. By σ^{r-1}, we denote the submatching of agents matched before round r. If $\sigma^{r-1} \in \overline{\mathcal{M}}$, then the algorithm proceeds with the following three steps of round r:
>
> *Step 1. Pointing.* Each house $h \in \overline{H_{\sigma^{r-1}}}$ points to the agent who controls it at σ^{r-1}. Each agent $i \in \overline{I_{\sigma^{r-1}}}$ points to her most preferred outcome in $\overline{H_{\sigma^{r-1}}}$.
>
> *Step 2(a). Matching Simple Trading Cycles.* A cycle
>
> $$h^1 \to i^1 \to h^2 \to ...h^n \to i^n \to h^1,$$

in which $n \in \{1, 2, ...\}$ and agents $i^\ell \in \overline{I_{\sigma^{r-1}}}$ point to houses $h^{\ell+1} \in \overline{H_{\sigma^{r-1}}}$ and houses h^ℓ point to agents i^ℓ (here $\ell = 1, ..., n$ and superscripts are added modulo n), is simple when one of the agents is an owner. Each agent in each simple trading cycle is matched with the house she is pointing to.

Step 2(b). Forcing Brokers to Downgrade Their Pointing. If there are no simple trading cycles in the preceding Step 2(a), and only then we proceed as follows (otherwise we proceed to step 3).

- If there is a cycle in which a broker i points to a brokered house, and there is another broker or owner that points to this house, then we force broker i to point to her next choice and we return to Step 2(a).[18]
- Otherwise, we clear all trading cycles by matching each agent in each cycle with the house she is pointing to.

Step 3. Submatching σ^r is defined as the union of σ^{r-1} and the set of newly matched agent-house pairs. When all agents or all houses are matched under σ^r, then the algorithm terminates and gives matching σ^r as its outcome.

One important feature of the TC mechanisms is that we can, without loss of generality, rule out the existence of brokers at some submatching σ if there is a single owner at σ. We formalize this property as a remark:

Remark 1. *Pycia and Ünver (2017) For each TC mechanism such that for some σ there is only one owner and one broker, there is an equivalent TC mechanism such that at σ there are no brokers and the same owner owns all houses.*

Using Theorem 1 and Pycia and Ünver (2017)'s characterization we obtain the following corollary:

Corollary 1. *A mechanism is individually strategy-proof and Arrovian efficient if and only if it is a TC mechanism.*

4 Complete Social Welfare Functions

So far, we allowed welfare functions to incompletely rank social outcomes. We now show that a class that we refer to as almost sequential dictatorships is exactly the mechanisms

[18] Importantly, broker i is unique by R1.

that are strategy-proof and Arrovian efficient with respect to complete SWF, that is SWF that always rank all outcomes.

First we define the following class: a **top-trading-cycles** (**TTC**) (or **hierarchical exchange**) mechanism is a TC mechanism with a control-rights structure in which no house is ever brokered at any submatching (Pápai, 2000). A TTC mechanism $\psi^{c,b}$ will be denoted by dropping b from its notation as ψ^{c}.

TTC mechanisms form a strict subclass of TC mechanisms. Let us start with an example showing that not every TTC is efficient with respect to a complete SWF.

Example 3. When $|H| > |I| = 2$, an agent cannot own two houses while a second agent owns a house: Consider allocating three houses to two agents. Let φ be a TTC mechanism in which agent 1 owns house A and agent 2 owns houses B and C. We will show that there is no complete SWF such that φ is efficient.

Consider the preference profile

$$\succ = \begin{array}{|c|c|} 1 & 2 \\ \hline B & A \\ A & B \\ C & C \end{array}.$$

Consider also the following four additional preference profiles

$$\succ^1 = \begin{array}{|c|c|} 1 & 2 \\ \hline B & C \\ A & \vdots \\ \vdots & \end{array}, \quad \succ^2 = \begin{array}{|c|c|} 1 & 2 \\ \hline B & B \\ C & C \\ \vdots & \vdots \end{array}, \quad \succ^3 = \begin{array}{|c|c|} 1 & 2 \\ \hline C & A \\ \vdots & B \\ & \vdots \end{array}, \quad \succ^4 = \begin{array}{|c|c|} 1 & 2 \\ \hline A & A \\ C & C \\ \vdots & \vdots \end{array}.$$

Denote

$$\begin{aligned} \mu^1 &= \varphi[\succ^1] = \{(1, B), (2, C)\}, \\ \mu^2 &= \varphi[\succ^2] = \{(1, C), (2, B)\}, \\ \mu^3 &= \varphi[\succ^3] = \{(1, C), (2, A)\}, \\ \mu^4 &= \varphi[\succ^4] = \{(1, A), (2, C)\}. \end{aligned}$$

Now, if there is a complete SWF Φ such that φ is Arrovian efficient, then $\Phi(\succ^1)$ ranks μ^1 above μ^4, and by IIA, this implies that $\Phi(\succ)$ ranks μ^1 above μ^4. Similarly, $\Phi(\succ^2)$ ranks μ^2 above μ^1, and by IIA, this implies that $\Phi(\succ)$ ranks μ^2 above μ^1. Further, and again similarly, $\Phi(\succ^3)$ ranks μ^3 above μ^2, and by IIA, this implies that $\Phi(\succ)$ ranks μ^3 above μ^2. Finally, $\Phi(\succ^4)$ ranks μ^4 above μ^3, and by IIA, this implies that $\Phi(\succ)$ ranks μ^4 above μ^3.

But then $\Phi(\succ)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient.

Observe that this example relies on the existence of more houses than agents. We will use this example to prove a theorem for the case with $|H| > |I|$, and we will consider the case $|H| = |I|$ later. To do this, we introduce sequential dictatorships formally. A **sequential dictatorship** is a TTC mechanism ψ^c such that for every $\sigma \in \overline{\mathcal{M}}$ and $h, h' \in \overline{H_\sigma}$, $c_h(\sigma) = c_{h'}(\sigma)$, i.e., an unmatched agent owns all unmatched houses at σ. For notational convenience, we will represent each $c_h(\cdot)$ as $c(\cdot)$. Sequential dictatorships turn out to be the class of Arrovian-efficient and individually strategy-proof mechanisms for this case:

Theorem 2. *Suppose $|H| > |I|$. A mechanism is individually strategy-proof and Arrovian efficient with respect to a complete social welfare function if and only if it is a sequential dictatorship.*

Proof of Theorem 2. If $|I| = 1$, the theorem is trivially true. Suppose $|I| \geq 2$.

($\Longrightarrow$) Consider a mechanism φ that is individually strategy-proof and efficient with respect to a complete Arrovian welfare function. By Theorem 1 and Corollary 1, φ is a TC mechanism $\psi^{c,b}$.

Fix an arbitrary preference profile $\succ \in \mathbf{P}$. We claim that at any round r of the algorithm $\psi^{c,b}$, there is exactly one agent who controls all houses. We prove it in two steps. First, let us show that there cannot be two (or more) agents who each own a house. By way of contradiction, suppose that some agent 1 controls house A and some other agent 2 controls house B in round r.

Suppose σ is the submatching created by the TC algorithm for $\psi^{c,b}$ before round r at $\succ$. Fix house $C \in \{A, B\}$ as an unmatched house at σ. Consider four auxiliary preference profiles $\succ^\ell$ that all share the following properties: (i) each agent matched under σ ranks houses under $\succ^\ell$, $\ell = 1, ..., 4$, in the same way they rank them under $\succ$, (ii) each agent i unmatched at σ and different from agents 1 and 2 ranks a unique σ-unmatched house $h_i \notin \{A, B, C\} \cup H_\sigma$ as her first choice (such a unique house exists as $|H| > |I|$), and (iii) agents 1 and 2 each rank all houses other than A, B, C lower than A, B, C. In particular, the four profiles differ only in how agents 1 and 2 rank houses A, B, C: the ranking of A, B, C is the same as in the four preference profiles of Example 3 above. Notice that

$$\psi^{c,b}[\succ^\ell] = \sigma \cup \mu^\ell \cup \{(i, h_i)\}_{i \in \overline{I_\sigma} - \{1,2\}},$$

where μ^ℓs are defined as in Example 3 above. Furthermore, the same argument we used in the example shows that there can be no SWF that ranks all four μ^ℓs, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{c,b}$ efficient, a contradiction that implies that there cannot be two agents who own houses in a round of the algorithm.

As $\psi^{c,b}$ never allows two owners in a round of the algorithm, by Corollary 1 and Remark 1, there are no brokers in any round, either. Hence, in each round of the algorithm there is a single agent who controls (and owns) all houses. That means that $\psi^{c,b}$ is a sequential dictatorship.

($\Longleftarrow$) Consider a sequential dictatorship ψ^{c}. We construct a complete SWF Φ such that ψ^{c} is efficient with respect to Φ. Under Φ any two matchings are ranked according to preferences of the first-round dictator; if she is indifferent , then the matchings are ranked according to the preferences of the second-round dictator, etc. Formally, for any $\succ \in \mathbf{P}$ and any two distinct $\mu, \nu \in \mathcal{M}$, let $\mu \Phi(\succ) \nu$ if and only if there exists $k \in \{1, ..., |I|\}$ such that $\mu(i_1) = \nu(i_1)$, ... and $\mu(i_{k-1}) = \nu(i_{k-1})$, and agent i_k prefers $\mu(i_k)$ over $\nu(i_k)$, where agents $i_1, ..., i_k$ are defined recursively: $i_1 = c(\emptyset)$, and in general $i_\ell = c(\{(i_1, \mu(i_1)), ..., (i_{\ell-1}, \mu(i_{\ell-1}))\})$ for $\ell = 1, ..., k$. It is straightforward to verify that Φ is a complete SWF and that ψ^{c} is efficient with respect to Φ. QED

Next we turn our attention to what happens when $|H| = |I|$. The above argument relies on the fact that there exists one extra house that can be used to regulate the ownership of all houses in any round of the algorithm. Suppose $|H| = |I|$. Then we can modify the argument in the proof an obtain a slightly different result. For this purpose we introduce a new class of mechanisms slightly larger than sequential dictatorships.

An **almost sequential dictatorship** is a TTC mechanism ψ^{c} such that for every $\sigma \in \overline{\mathcal{M}}$ such that $|\overline{H_\sigma}| \neq 2$ we have $c_h(\sigma) = c_{h'}(\sigma)$ for every $h, h' \in \overline{H_\sigma}$.

Therefore, the only mechanisms that are not sequential dictatorships in this class are mechanisms that assign to different owners each of the houses when only two houses (and hence, two agents) are left, but otherwise a single agent owns all houses.

Our third result is as follows:

Theorem 3. *A mechanism is individually strategy-proof and Arrovian efficient with respect to a complete SWF if and only if it is an almost sequential dictatorship.*

First, we modify Example 3 and show that an agent cannot own multiple houses while one other agent owns a house, and then we show in two examples that three agents each cannot simultaneously control a house under a TC mechanism that is efficient with respect to a complete SWF. We will use these three examples in proving Theorem 3.

Example 4. When $|H| = |I| = 3$, an agent cannot own two houses while another agent owns the third house: Let φ be a TTC mechanism in which agent 1 owns house

A, agent 2 owns houses B and C, and hence, agent 3 does not control any house. Consider 5 preference profiles $\succ, \succ^1, \succ^2, \succ^3, \succ^4$, as in Example 3. Suppose the preferences of agents 1 and 2 are exactly the same as in Example 3 under the respective profiles, while agent 3 has the same arbitrarily fixed preference relation $\succ_3 = \succ_3^1 = \ldots = \succ_3^4$. Denote

$$\mu^1 = \varphi[\succ^1] = \{(1, B), (2, C), (3, A)\},$$
$$\mu^2 = \varphi[\succ^2] = \{(1, C), (2, B), (3, A)\},$$
$$\mu^3 = \varphi[\succ^3] = \{(1, C), (2, A), (3, B)\},$$
$$\mu^4 = \varphi[\succ^4] = \{(1, A), (2, C), (3, B)\}.$$

Using the exact same argument as in Example 3, we establish that $\Phi(\succ)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient.

Example 5. When $|H| = |I| = 3$, one agent cannot control a house while the others each own a house: Let φ be a TTC mechanism in which agent 1 owns house A, agent 2 owns house B, and agent 3 controls house C. We will show that there is no complete SWF such that φ is Arrovian efficient.

Consider the preference profile

$$\succ = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & C & A \\ C & A & B \\ A & B & C \end{array}.$$

Consider also the following three additional preference profiles

$$\succ^1 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & C & B \\ C & \vdots & \vdots \\ A & & \end{array}, \quad \succ^2 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline C & C & A \\ \vdots & A & \vdots \\ & B & \end{array}, \quad \succ^3 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & A & A \\ \vdots & \vdots & B \\ & & C \end{array}.$$

Regardless of whether agent 3 owns or brokers house C, we have

$$\mu^1 = \varphi[\succ^1] = \{(1, A), (2, C), (3, B)\};$$
$$\mu^2 = \varphi[\succ^2] = \{(1, C), (2, B), (3, A)\};$$
$$\mu^3 = \varphi[\succ^3] = \{(1, B), (2, A), (3, C)\}.$$

If there is a complete SWF Φ such that φ is Arrovian efficient, then $\Phi(\succ^1)$ ranks μ^1 above μ^3, and by IIA, this implies that $\Phi(\succ)$ ranks μ^1 above μ^3. Similarly, $\Phi(\succ^2)$ ranks μ^2 above μ^1, and by IIA, this implies that $\Phi(\succ)$ ranks μ^2 above μ^1. Further, and again similarly, $\Phi(\succ^3)$ ranks μ^3 above μ^2, and by IIA, this implies that $\Phi(\succ)$ ranks μ^3 above μ^2. Then $\Phi(\succ)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient.

Example 6. When $|H| = |I| = 3$, there cannot be three brokers: Let φ be a TTC mechanism in which agent 1 brokers house A, agent 2 brokers house B, and agent 3 brokers house C. We will show that there is no complete SWF such that φ is Arrovian efficient.

$$\succ = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & B & C \\ A & A & B \\ C & C & A \end{array}.$$

Consider also the following three additional preference profiles

$$\succ^1 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & B & C \\ C & A & B \\ \vdots & \vdots & \vdots \end{array}, \quad \succ^2 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & B & C \\ A & C & A \\ \vdots & \vdots & \vdots \end{array}, \quad \succ^3 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & A & B \\ C & C & A \\ \vdots & \vdots & \vdots \end{array}.$$

Denote

$$\begin{aligned} \mu^1 &= \varphi[\succ^1] = \{(1, A), (2, B), (3, C)\}; \\ \mu^2 &= \varphi[\succ^2] = \{(1, B), (2, C), (3, A)\}; \\ \mu^3 &= \varphi[\succ^3] = \{(1, C), (2, A), (3, B)\}. \end{aligned}$$

If there is a complete SWF Φ such that φ is Arrovian efficient, then $\Phi(\succ^1)$ ranks μ^1 above μ^3, and by IIA, this implies that $\Phi(\succ)$ ranks μ^1 above μ^3. Similarly, $\Phi(\succ^2)$ ranks μ^2 above μ^1, and by IIA, this implies that $\Phi(\succ)$ ranks μ^2 above μ^1. Further, again similarly, $\Phi(\succ^3)$ ranks μ^3 above μ^2, and by IIA, this implies that $\Phi(\succ)$ ranks μ^3 above μ^2. Then $\Phi(\succ)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient.

Proof of Theorem 3. If $|H| > |I|$, it follows from Theorem 2. So suppose $|H| = |I|$. If $|I| = 1$, the theorem is trivially true. So suppose $|I| > 1$:

($\Longrightarrow$) Consider a mechanism φ that is individually strategy-proof and efficient with respect to a complete Arrovian welfare function. By Theorem 1 and Corollary 1, φ is a TC mechanism $\psi^{c,b}$.

Fix $\succ \in \mathbf{P}$. We claim that at any round r of the algorithm for $\psi^{c,b}$, there is exactly one agent who controls all houses whenever $|\overline{I_\sigma}| > 2$. We prove it in three steps (in accordance with Examples 4-6). Let σ be the submatching created by the algorithm $\psi^{c,b}$ before round r for $\succ$.

- First, we show that an agent cannot own two houses while another agent owns a third house: By way of contradiction, suppose that some agent 1 owns house A and agent 2 owns houses B and C in round r. Then there exists an agent 3 who does not control any house at round r as $|H| = |I|$. Consider four auxiliary preference profiles $\succ^\ell$ that all share the following properties: (i) each agent matched under σ ranks houses under $\succ^\ell$, $\ell = 1, ..., 4$, in the same way they rank them under $\succ$, (ii) each agent i unmatched at σ and different from agents $1, 2, 3$ ranks a unique σ-unmatched house $h_i \notin \{A, B, C\} \cup H_\sigma$ as her first choice (such a unique house exists as $|H| = |I|$), (iii) agents 1 and 2 each rank all houses other than A, B, C lower than A, B, C, and (iv) agent 3's preferences are the same as $\succ_i$ under all four profiles. In particular, the four profiles differ only in how agents 1 and 2 rank houses A, B, C: the ranking of A, B, C is the same as in the four preference profiles of Example 4 above. Notice that

$$\psi^{c,b}[\succ^\ell] = \sigma \cup \mu^\ell \cup \{(i, h_i)\}_{i \in \overline{I_\sigma} - \{1,2,3\}},$$

where μ^ℓs are defined as in Example 4 above. Furthermore, the same argument we used in Example 4 shows that there can be no SWF that ranks all four μ^ℓs, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{c,b}$ efficient, a contradiction.

- Next, we show that one agent cannot control a house while at least two others each own a house in round r: Suppose, to the contrary, agent 1 owns house A, agent 2 owns house B, and agent 3 controls house C in round r. Consider three auxiliary preference profiles $\succ^\ell$ that all share the following properties: (i) each agent matched under σ ranks houses under $\succ^\ell$, $\ell = 1, 2, 3$, in the same way they rank them under $\succ$, (ii) each agent i unmatched at σ and different from agents $1, 2, 3$ ranks a unique σ-unmatched house $h_i \notin \{A, B, C\} \cup H_\sigma$ as her first choice (such a unique house exists as $|H| = |I|$), and (iii) agents 1,2, 3 each rank all houses other than A, B, C lower than A, B, C, and the ranking of A, B, C is the same as in the three preference profiles of Example 5 above. Observe that

$$\psi^{c,b}[\succ^\ell] = \sigma \cup \mu^\ell \cup \{(i, h_i)\}_{i \in \overline{I_\sigma} - \{1,2,3\}},$$

where μ^ℓs are defined as in Example 5 above. Furthermore, the same argument we used in Example 5 shows that there can be no SWF that ranks all three μ^ℓs, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{c,b}$ efficient, a contradiction.

- Finally, using a variant of Example 6, we show that there cannot be multiple brokers at round r (as multiple brokers can only occur with 3 agents and 3 houses, where each agent brokers a distinct house): Suppose not. Then consider three auxiliary preference profiles $\succ^\ell$ that all share the following properties: (i) each agent matched under σ ranks houses under $\succ^\ell$, $\ell = 1, 2, 3$, in the same way they rank them under $\succ$, (ii) agents 1,2, 3, who are the only remaining unmatched agents, each rank all houses other than A, B, C lower than A, B, C, and (iii) the ranking of A, B, C is the same as in the three preference profiles of Example 6 above. Notice that

$$\psi^{c,b}[\succ^\ell] = \sigma \cup \mu^\ell,$$

where μ^ℓs are defined as in Example 6 above. Furthermore, the same argument we used in Example 6 shows that there can be no SWF that ranks all three μ^ℓs, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{c,b}$ efficient, a contradiction.

Thus, a single agent owns all houses at round r when σ is fixed for $|\overline{I_\sigma}| > 2$ (by Corollary 1 and Remark 1).

This means that $\psi^{c,b}$ is an almost sequential dictatorship, as all TC mechanisms restricted to only two agents are almost sequential dictatorships.

($\Longleftarrow$) Consider an almost sequential dictatorship ψ^c. If ψ^c is a sequential dictatorship, then the proof of Theorem 2 works. So suppose it is not a sequential dictatorship. Hence, $|H| = |I|$. We construct a complete SWF Φ such that ψ^c is efficient with respect to Φ. Under Φ any two matchings are ranked according to preferences of the first-round dictator; if she is indifferent , then the matchings are ranked according to the preferences of the second-round dictator, etc., until only two agents remain unmatched. At this round let 1 and 2 be the two agents and A and B be the two houses remaining unmatched. Observe that there are only two matchings, μ and ν, in which all agents' assignments are the same but the last two: in one 1 gets A and 2 gets B, and in the other vice versa. Then one of these two matchings is equal to $\psi^c[\succ']$, where $\succ'$ ranks the assignment of any agent other than 1 and 2 in μ (or equivalently ν) as her first choice, and for 1 and 2, the new preferences are the same as the original ones under $\succ$. We rank $\psi^c[\succ'] \in \{\mu, \nu\}$ before the other one under $\Phi(\succ)$.

Formally, for every $\mu \in \mathcal{M}$, let sequential dictators $i_1, \ldots, i_{|I|-2}$ be defined as $i_1 = c_h(\emptyset)$ for every $h \in H$, and in general, $i_\ell = c_h(\{(i_1, \mu(i_1)), ..., (i_{\ell-1}, \mu(i_{\ell-1}))\})$ for every $h \in$

$H - \{\mu(i_1), ...\mu(i_{\ell-1})\}$ and $\ell = 1, ..., k$; then for every $\nu \in \mathcal{M} - \{\mu\}$, we say $\mu\, \Phi(\succ)\, \nu$ if one of the following two conditions holds:

1. there exists $k \in \{1, ..., |I| - 2\}$ such that $\mu(i_1) = \nu(i_1)$, ..., $\mu(i_{k-1}) = \nu(i_{k-1})$, and $\mu(i_k) \succ_{i_k} \nu(i_k)$;

   or

2. for every $\ell \in \{1, ..., |I| - 2\}$, $\mu(i_\ell) = \nu(i_\ell)$, and for $\succ' \in \mathbf{P}$ where each i_ℓ ranks $\mu(i_\ell)$ first while the remaining two agents have the same preferences as in $\succ$, we have $\psi^c[\succ'] = \mu$.

By construction, Φ is complete, antisymmetric, and transitive. Moreover, it satisfies the Pareto postulate. To see that it also satisfies IIA, consider two distinct matchings, μ and $\nu \in \mathcal{M}$, and $\succ \in \mathbf{P}$ such that $\mu\, \Phi(\succ)\, \nu$. Also consider another profile $\hat{\succ} \in \mathbf{P}$ such that each agent i's preference over the two matching assignments is the same in $\hat{\succ}_i$ as in $\succ_i$. If $\mu\, \Phi(\succ)\, \nu$ because of condition 1 above, then condition 1 continues to hold for $\hat{\succ}$ and thus $\mu\, \Phi(\hat{\succ})\, \nu$. On the other hand, if $\mu\, \Phi(\succ)\, \nu$ because of condition 2 above, then μ and ν only differ in how the last two agents are assigned the remaining two houses. Hence, the profile constructed to check condition 2 for $\mu\, \Phi(\hat{\succ})\, \nu$, which we refer to as $\hat{\succ}'$, would lead to $\psi^c[\hat{\succ}'] = \mu$ because:

1. the first $|I| - 2$ dictators would still get their μ assignments in the first $|I \vdash 2$ rounds of the TC algorithm for $\psi^c[\hat{\succ}']$, and

2. thus, the assignment of remaining two agents under $\psi^c[\hat{\succ}']$ would be identical with that under μ as the relative ranking of their assignments under μ and ν are identical both in $\succ$ and $\hat{\succ}$, and the ranking of the other houses do not matter for finding the outcome of the almost serial dictatorship.

Thus, $\mu\, \Phi(\hat{\succ})\, \nu$, showing Φ satisfies IIA. QED

References

Abdulkadiroğlu, A., Y.-K. Che, P. A. Pathak, A. E. Roth, and O. Tercieux (2020): "Efficiency, Justified Envy, and Incentives in Priority-Based Matching," *American Economic Review: Insights*, forthcoming.

Abdulkadiroğlu, A., and T. Sönmez (1998): "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems," *Econometrica*, 66, 689–701.

——— (1999): "House Allocation with Existing Tenants," *Journal of Economic Theory*, 88, 233–260.

——— (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Akbarpour, M., and S. Li (2020): "Credible Auctions: A Trilemma," *Econometrica*, 88 (2), 425–467.

Arrow, K. J. (1963): *Social Choice and Individual Values*. New York: Wiley, 2nd edn.

Austen-Smith, D., and J. S. Banks (1999): *Positive Political Theory I: Collective Preference*. University of Michigan Press.

Barberà, S. (1983): "Strategy–Proofness and Pivotal Voters: A Direct Proof of the Gibbard–Satterthwaite Theorem," *International Economic Review*, 24, 413—417.

Barberà, S., and M. O. Jackson (1995): "Strategy-proof Exchange," *Econometrica*, 63, 51–87.

Bergson, A. (1938): "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, 52(2), 310–334.

Black, D. (1948): "On the Rationale of Group Decision-making," *Journal of Political Economy*, 56(1), 23–34.

Blair, D., and E. Muller (1983): "Essential Aggregation Procedures on Restricted Domains of Preferences," *Journal of Economic Theory*, 30, 34–53.

Bogolomania, A., and H. Moulin (2004): "Random Matching Under Dichotomous Preferences," *Econometrica*, 72, 257–279.

Bogomolnaia, A., and H. Moulin (2001): "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 100, 295–328.

Campbell, D. E., and J. S. Kelly (2003): "A Strategy-Proofness Characterization of Majority Rule," *Economic Theory*, 22(3), 557–568.

Dasgupta, P., P. Hammond, and E. Maskin (1979): "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 46, 185–216.

Ehlers, L. (2002): "Coalitional Strategy-Proof House Allocation," *Journal of Economic Theory*, 105, 298–317.

EHLERS, L., AND B. KLAUS (2003): "Resource-Monotonicity for House Allocation Problems," *International Journal of Game Theory*, 32, 545–560.

EHLERS, L., B. KLAUS, AND S. PÁPAI (2002): "Strategy-Proofness and Population-Monotonicity for House Allocation Problems," *Journal of Mathematical Economics*, 38, 329–339.

ERGIN, H. (2000): "Consistency in House Allocation Problems," *Journal of Mathematical Economics*, 34, 77–97.

GIBBARD, A. (1973): "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41, 587–601.

HAKIMOV, R., AND O. KESTEN (2014): "The equitable top trading cycles mechanism for school choice," Working paper.

HATFIELD, J. W. (2009): "Strategy-Proof, Efficient, and Nonbossy Quota Allocations," *Social Choice and Welfare*, 33 No. 3, 505–515.

HYLLAND, A., AND R. ZECKHAUSER (1979): "The Efficient Allocation of Individuals to Positions," *Journal of Political Economy*, 87, 293–314.

JARAMILLO, P., AND V. MANJUNATH (2012): "The difference indifference makes in strategy-proof allocation of objects," *Journal of Economic Theory*, 147(5), 1913–1946.

KARAKAYA, M., B. KLAUS, AND J. C. SCHLEGEL (2019): "Top trading cycles, consistency, and acyclic priorities for house allocation with existing tenants," *Journal of Economic Theory*, 184, 104948.

KLAUS, B., AND E. MIYAGAWA (2002): "Strategy-proofness, solidarity, and consistency for multiple assignment problems," *International Journal of Game Theory*, 30(3), 421–435.

LI, S. (2015): "Obviously Strategy-Proof Mechanisms," Working Paper.

LIU, Q., AND M. PYCIA (2011): "Ordinal Efficiency, Fairness, and Incentives in Large Markets," Working Paper.

MA, J. (1994): "Strategy-proofness and the strict core in a market with indivisibilities," *International Journal of Game Theory*, 23, 75–83.

MASKIN, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

Morrill, T. (2013): "An Alternative Characterization of Top Trading Cycles," *Economic Theory*, 54, 181–197.

——— (2015): "Two Simple Variations of Top Trading Cycles," *Economic Theory*, 60, 123–140.

Moulin, H. (1988): *Axioms of Cooperative Decision Making*. Cambridge: Cambridge University Press.

Pai, M. M., and A. Roth (2018): "Privacy and Mechanism Design," Working Paper.

Pápai, S. (2000): "Strategyproof Assignment by Hierarchical Exchange," *Econometrica*, 68, 1403–1433.

——— (2001): "Strategyproof and Nonbossy Multiple Assignments," *Journal of Public Economic Theory*, 3, 257–271.

——— (2007): "Exchange in a General Market with Indivisible Goods," *Journal of Economic Theory*, 132, 208–235.

Pycia, M. (2016): "Swaps on Networks," Working Paper.

——— (2019): "Evaluating with Statistics: Which Outcome Measures Differentiate Among Matching Mechanisms?," Working paper.

Pycia, M., and P. Troyan (2019): "A Theory of Simplicity on Games and Mechanism Design," Working paper.

Pycia, M., and M. U. Ünver (2007): "Outside Options in Neutral Discrete Resource Allocation," Working paper.

——— (2011): "Trading Cycles for School Choice," Working Paper Available at SSRN: http://dx.doi.org/10.2139/ssrn.1899344.

——— (2017): "Incentive Compatible Allocation and Exchange of Discrete Resources," *Theoretical Economics*, 12, 287–329.

Root, J., and D. Ahn (2020): "Incentives and Efficiency in Constrained Allocation Mechanisms," Working Paper.

Roth, A. E., T. Sönmez, and M. U. Ünver (2004): "Kidney Exchange," *Quarterly Journal of Economics*, 119, 457–488.

Samuelson, P. A. (1947): *Foundations of Economic Analysis*. MA: Harvard University Press.

Satterthwaite, M. (1975): "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 10, 187–216.

Satterthwaite, M., and H. Sonnenschein (1981): "Strategy-Proof Allocation Mechanisms at Differentiable Points," *Review of Economic Studies*, 48, 587–597.

Sen, A. (1970): "Interpersonal Aggregation and Partial Comparability," *Econometrica*, 38(3), 393–409.

——— (1999): "The Possibility of Social Choice," *American Economic Review*, 89(3), 349–378.

Shapley, L., and H. Scarf (1974): "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1, 23–37.

Sönmez, T., and M. U. Ünver (2010): "House Allocation with Existing Tenants: A Characterization," *Games and Economic Behavior*, 69, 425–445.

Svensson, L.-G. (1994): "Queue Allocation of Indivisible Goods," *Social Choice and Welfare*, 11, 223–230.

——— (1999): "Strategy-proof Allocation of Indivisible Goods," *Social Choice and Welfare*, 16, 557–567.

Takamiya, K. (2001): "Coalition Strategy-Proofness and Monotonicity in Shapley-Scarf Housing Markets," *Mathematical Social Sciences*, 41, 201–213.

Tang, Q., and Y. Zhang (2015): "Hierarchical Exchange Rules and the Core in Indivisible Objects Allocation," .

Weymark, J. (1984): "Arrow's theorem with social quasi-orderings," *Public Choice*, 42, 235–246.

Woodward, K. (2020): "Self-Auditable Auctions," Working paper.

Zhou, L. (1991): "Inefficiency of Strategy-Proof Allocation Mechanisms in Pure Exchange Economies," *Social Choice and Welfare*, 8, 247–257.

ZWICKER, W. S. (2016): *Introduction to the Theory of Voting*pp. 23–56. Cambridge University Press.

A An Incomplete Arrovian Social Welfare Function

The following example illustrates an incomplete Arrovian SWF.

Example 7. Consider a society (or an employer) assigning one task to each of three employees. All the tasks need to be completed, and the society would like to respect the preferences of the employees in assigning the tasks as much as possible. As a second order concern, the society would like to avoid assigning Task A to employee 1 (e.g. because of a belief that employee 1 is not very good in doing this job). The society thus has an SWF that has the maximum at a Pareto-efficient matching that does not assign Task A to employee 1 if there exists at least one Pareto-efficient matching that does not assign Task A to employee 1.

The society's SWF can be equivalently described in terms of a Trading Cycles mechanism ψ in which employee 1 brokers A, employee 2 has ownership of B and employee 3 has ownership of C: for any preference profile $\succ_{\{1,2,3\}}$, the SWF $\Psi(\succ)$ ranks any two distinct matchings μ and ν if and only if $\mu = \psi\,[\succ]$ or μ Pareto dominates ν; the social ranking is then $\mu\ \Psi(\succ)\ \nu$.

For instance, for the preference profile

$$\succ = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & A & B \\ B & B & C \\ C & C & A \end{array},$$

the outcome of Trading Cycles ψ is $\psi[\succ] = \{(1, B)\,, (2, A)\,, (3, C)\}$, and the ranking of the matchings with respect to $\Psi(\succ)$ is given in Figure 1.

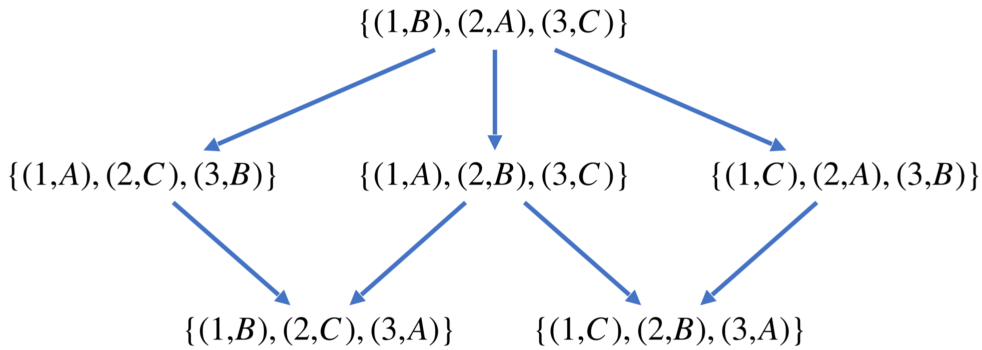

Figure 1: $\Psi(\succ)$ in Example 7. For matching μ, ν, we have $\mu \; \Psi(\succ) \; \nu$ if and only if there is a directed path from μ to ν in this graph.